A Greedy-Based Approach of Fast Transaction Broadcasting in Bitcoin Networks

ABSTRACT

In this paper, we premeditate the problem of designing a model of information propagation with solving the bifurcation problem of the blockchain system. We define a new diffusion mode with spreading information through the designated seed nodes in the network for the whole system which selects the nodes based on a specific problem. Therefore, we derive a novel influence time minimization (ITM) problem which is that how to choose a group of seed nodes as the source of information dissemination process so that the time required for the entire network to be infected is the smallest. We prove that the problem is NP-hard. A greedy-based algorithm is also proposed. We further propose a provable approximation guarantee for the solution of the algorithm based on the analysis of the monotone submodular function.

Experiment results show that the new propagation mode defined by this work can indeed significantly reduce the propagation time in Bitcoin network compared with the traditional method, which further illustrate that greedy-based algorithm can provide a solution with better performance for ITM problem. Moreover, we find the new program has certain practical significance for improving not only the speed of information propagation but operating efficiency of Bitcoin network.

KEYWORDS

Bitcoin network, information propagation, greedy-algorithm, approximation guarantee

ACM Reference Format:

. 2019. A Greedy-Based Approach of Fast Transaction Broadcasting in Bitcoin Networks. In *Proceedings of ACM TURC 2019 (TURC2019)*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

As a decentralized, non-hierarchical system , the blockchain can be defined as a decentralized, unalterable distributed ledger. Its update and confidentiality depend on the distributed structure between peer-to-peer (P2P) network users. There are some inherent defects in the structure of the blockchain, one of the most important problems is bifurcation problem, what is the inconsistency of the blockchain in Bitcoin network caused by the inconsistency of newly generated block connections, due to the uncertainty of information sources, the delay of information propagation and the diversity of network structures. The process of delivering transactions and blocks is realized by broadcasting from sources who find a new block or want to do a transaction to the other users in the network. The consistency protocol in Bitcoin network is proposed to solve such problems. However, some malicious users in Bitcoin network will exploit some of the weaknesses in the blockchain technology to perform some malicious attacks. For example, fork attack, bifurcation attack by bribery, self-propelled mining, etc.

Most of existing literature emphasized the process of source backtracking in Bitcoin network to track criminals who trade illegal items by Bitcoin in real world. In this paper, we focus on the design of transaction model to eliminate the bifurcation phenomenon and accelerate the information propagation in Bitcoin network so that improve the efficiency of Bitcoin system. We consider a method to spreading information to the whole network by selecting a group of seed nodes. We further think about how to choose the seed nodes. Moreover, we develop plans against the property of the problem and implement the research programs.

The main contributions of our work are as follows:

(1) We first consider a new information propagation mode which spread to the whole network by choosing some special nodes as source nodes. The method introduces partial centrality in some aspects, but it doesnt undermine the essential distribute attribute of Bitcoin network which we will explain in detail.

(2) We first derive the ITM problem in Bitcoin network for the model designing of information propagation. We further prove the problem is NP-hard.

(3) We propose a greedy-based method for the ITM problem, and derive the approximate guarantee of the algorithm result. We prove the approximate relationship by illustrating the objective function of ITM problem is a monotonic submodule function.

Organization. The rest of the paper is organized as follows. In Section 2, we review some related works. In the Section 3, we introduce a spreading model which is widely used and formulate the problem we derived. In Section 4, we propose a greedy-based algorithm and the approximate guarantee. In Section 5, we show and analyze the simulation results. In Section 6, we conclude our work.

2 RELATED WORK

Bitcoin transaction security is a primary goal that measures whether the system is effective. Therefore, anonymity has

TURC2019, MAY 18, 2019, Chengdu, China

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

been widely studied and discussed in recent years as an important criterion for security measurement in Bitcoin network [1]. The spread of transaction information as one of the most important parts with the function of conducting new blocks and pushing on the process of transactions has gradually been valued by people. At present, the main research contents in the Bitcoin network are as follows. Analysis of transaction anonymity and transaction backtracking under different communication protocols [2], design and exploration of information dissemination models in networks under consideration of anonymity [3], investigate the problem of information dissemination in different scenarios [4], design of the mechanism [5], protocol, architecture of Bitcoin network. In order to solve the problem of blockchain bifurcation and improve operational efficiency, this paper studies the propagation mode of Bitcoin network.

3 SYSTEM MODEL AND PROBLEM FORMULATION

3.1 Trickle Spreading Protocol

Trickle spreading protocol is a distribute flooding protocol. We can simplify the process of information propagation in Bitcoin network as a discrete time process. The main content of this protocol can be briefly stated as follows. In every time stamp, each node randomly selects an adjacent node as its communication object to infect. The distributed feature of the protocol is reflected in the randomness of each user to select neighbors, that is all the selected neighbors of the same user can form a queue which is random in order. It should be noted that during the propagation of a message, once a node is infected, its state does not change during the propagation. When all users in the network are notified, the propagation of a message is completed.

3.2 ITM Problem

In traditional blockchain protocols, the user who want to post the information broadcasts as the unique source to all users in the network. In order to solve the fork problem, we propose a new method of information dissemination to distinguish it from the traditional broadcast mode without changing the information communication protocol. In the newly proposed information dissemination mode, once a user who plans to publish information on the network, the user can transmit the information to a specific set of users instead of directly broadcasting to the entire network by himself. A user set (seed set) broadcasts as an information source set to other users. The seed users are selected depend on the time taken when each user receives the corresponding information. We select the users who spend the least amount of time. Therefore, we propose a new Influence Time Minimization problem (ITM problem) for the blockchain scenario:

Given a n-nodes directed graph $G = (V, E)$ and an integer k, we are aiming to select a set S with k-nodes as the information propagation source in graph G so that each node in Bitcoin network can be notified in a minimum amount of time. Which is formulated formally as follows.

$$\operatorname{argmin}_{S \subseteq V} E\left[T\left(S\right)\right] \tag{1}$$

$$I\left(S\right) = V, |S| \leq k \tag{2}$$

$$E\left[T\left(S\right)\right] = \sum\nolimits_{X \in \Omega} \Pr\left[X\right] \cdot T\left(S \,|X\right) \tag{3}$$

We use graph G to represent Bitcoin network, the n nodes in the graph represent n users in Bitcoin network. V represents the set of users in Bitcoin network. E represents the relationship between users. The k is the expected size of the selected seed set. S is the objective seed set to be selected. This paper does not distinguish the relationship between graph G and Bitcoin network, the relationship between nodes and users, the relationship between edges in G and connections with users. $I(S)$ represents the set of nodes that are notified at a certain time during the information propagation with seed set S. $T(S)$ represents the final time (starting time is 0) with the information propagation through Bitcoin network completed with source S. Due to the randomness of the process of information dissemination, the results with the same source S is not certain. Therefore, we use the expectation of $T(S)$ to express the spreading ability against a set of nodes. We utilize Ω to represent the sampling space of a complete propagation process with X on behalf of one sample in the sampling space and $Pr[X]$ representing the probability of occurrence for the sample X.

For the ITM problem mentioned above, we first provide the proof of its NP-hardness to further propose a suitable approximation algorithm. We finish our proof by transforming [6] a NP-complete vertex cover problem [7] to the ITM problem. Given an undirected graph, a vertex cover problem can be formulated as follows.

$$\operatorname{argmin}_{S \subseteq V} |S| \tag{4}$$

$$\forall e \in E, \exists v_e \in S \tag{5}$$

Where S is the minimum set of nodes to be sought in the vertex cover problem. The sign e represents an edge in the graph G. represents an endpoint of the edge e. Definitions for G, E, V are the same as defined in Section 3.

Based on the objective network, we construct another network to derive the ITM problem. We will prove the vertex cover problem in the original network can be solved by handling the ITM problem in the new network. As an example we show the construction in Figure 1. However, for the network with any scale or structure, we can utilize the similar way to construct a useful network. The process of construction is as follows.

First, find the maximum degree in the graph and add an extra node between every two nodes which connect with each other. Finally, for the nodes whose degree is less than the maximum degree, we add nodes for them in some new directions then using edges to construct connections between them. The last strategy focused on the nodes in the original graph, and each operating adds one node for a single one by an edge. The blue nodes shown in Figure 1 is original, the green one is constructed. After the process, the original nodes in constructed network have the same degree which is

the maximum value in the original graph. Bitcoin network is often modeled as directed graph, but the edges are two-direction connected which can be equivalent to undirected graph. The ITM problem defined in constructed network can

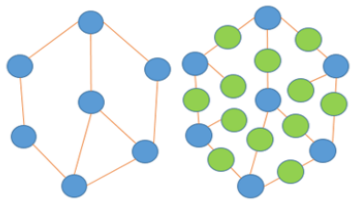

Figure 1: Left: Original Network. Right: Constructed Network.

be formulated as follow.

$$\text{argmin}_{S\subseteq V} E\left[T\left(S\right)\right] \tag{6}$$

$$I\left(S\right) = V, |S| \leq k, E\left[T\left(S\right)\right] \leq d+1 \tag{7}$$

$$E\left[T\left(S\right)\right] = \sum\nolimits_{X\in\Omega} \Pr\left[X\right] \cdot T\left(S\,|X\right) \tag{8}$$

Compare with the problem in Section 3. There is an addition constrain in 7. However, this constrain will take no effect to the property of the ITM problem. We can solve a series of problems to eliminate the affect from the constrain, what is when we set the d from 0 to a constant M there is always a minimum value N that satisfy the condition and permit the problem produce a feasible solution. The discussion above illustrates that we can settle the constructed ITM problem by solving the general ITM problem for N times. Therefore, the difficulty of the problem behaves linear relationship with general ITM problem. We can find from the two networks that once the ITM problem in constructed network is settled down, there is always a feasible solution for the vertex cover problem defined in the original network corresponding the solution of the ITM problem intuitively. Furthermore, the solution which violates the constrains can never be picked for the ITM problem. Through the process discussed above we finish the transformation from the vertex cover problem to the ITM problem. Therefore, the ITM problem is NP-hard.

4 ALGORITHM AND ANALYSIS

4.1 Greedy Algorithm

We use greedy algorithm [8] to solve the NP-hard ITM problem defined in Bitcoin network. In this paper, the main idea of the greedy algorithm is to traverse the different combinations of all nodes as seed sets in the network under different conditions for the process of information dissemination, continuously select the solution with the largest edge gain adding the node to the seed set until the requirements are met. The specific flow of the algorithm is shown in the figure below. S is the target seed set, F will be explained in detail later, which is the embodiment of the objective function T, what is the larger the value of F, the smaller the value of T, then the stronger the propagation ability of the corresponding set

$$\begin{array}{l} input\ S = \emptyset, V \\ for\ i = 1 \rightarrow k \\ \quad for\ \forall v \in V \cap S^c \\ \quad\quad calculate\ E[F(S \cup v)] \\ \quad end \\ \quad S \leftarrow S \cup \{argmax_v E[F(S \cup v)]\} \\ end \\ output\ S \end{array}$$

Figure 2: Greedy Algorithm

of seed nodes. k is the number of nodes required for S, v represents a node in the traversal process. V represents the set of nodes in the entire network.

At the beginning of the algorithm, the target seed set and the number of loops are initialized. S is initialized to an empty set. In the next step, the k-node target seed set is selected. We continuously select the node that maximize the edge gain of the objective function, that is trying to contain the remaining nodes in the network with the existing seed node set. Comparing the feasible remaining nodes, make sure that by expanding the existing seed set, the expected value of the objective function increases, and the node that maximizes the expectation of the objective function is added to the seed set. The process is looped until the number of elements in the target set reaches k. Then the algorithm ends, and the resulting set S is our target set. From the algorithm flowchart, we can conclude that the time complexity of the greedy algorithm in this paper is $O\left(k\,|V|\,R\right)$ by some preprocessing, where R is the cycle times. Preprocessing is the related calculation of the propagation process with different conditions, and no special consideration is taken in this paper [9].

4.2 Approximate Guarantee

In this section, we propose the approximate guarantee of the algorithm results, which is the upper bound of the time required to spread the information on the whole network.

We first construct a new objective function F and prove that F is a monotonic submodule function [10]. From this, we can get the approximate guarantee of the original objective function according to the new function. The function is constructed as follows.

$$F\left(S\right) = C - T\left(S\right) \geq 0 \tag{9}$$

We require F to be a non-negative function and C to be the upper bound of the function T. In bitcoin network, the upper bound of T is obviously exist, an upper bound is the number of nodes in the network. When the network structure is not conducive to information diffusion, the upper bound of the propagation time is $|V|$. From the expression, it can be concluded that F and T are in a negative linear relationship. Considering that T represents the propagation time, it is an indicator to measure the ability of the seed set to spread. The

smaller the T, the stronger the seed's ability to spread, and vice versa. So we can think of F as a function of measuring the ability of a set of seeds to be positively correlated. Hence, we will prove that the function F is a monotonic submodule function.

Considering the practical meaning of the function, if there are two sets that one contains the other one, let the two sets act the seed sets for the information propagation respectively. The spreading time the larger set needs must smaller than the subset. So T is a monotonic function apparently established. Hence, F is a monotonic function under the linear relationship with T. Furthermore, we prove F is a submodular function which satisfy the expressions as follows.

$$F\left(A_1 \cup \{v\}\right) - F\left(A_1\right) \geq F\left(A_2 \cup \{v\}\right) - F\left(A_2\right) \tag{10}$$

$$\Leftrightarrow T\left(A_1\right) - T\left(A_1 \cup \{v\}\right) \geq T\left(A_2\right) - T\left(A_2 \cup \{v\}\right) \tag{11}$$

Consider the relationship between time function T and influence function I and the monotonicity of T, we can transform the inequality.

$$\Leftrightarrow T\left(I_{t=T(A_1\cup\{v\})}\left(A_1\right)\right) \geq T\left(I_{t=T(A_2\cup\{v\})}\left(A_2\right)\right) \tag{12}$$

$$\Leftrightarrow I_{t=T(A_1\cup\{v\})}\left(A_1\right) \subseteq I_{t=T(A_2\cup\{v\})}\left(A_2\right) \tag{13}$$

$$\Leftrightarrow V \cap \left[I_{t=T(A_1\cup\{v\})}\left(A_1\right)\right]^C \supseteq V \cap \left[I_{t=T(A_2\cup\{v\})}\left(A_2\right)\right]^C \tag{14}$$

$$\Leftrightarrow I\left\{v \,|A_1 \cup \{v\}\right\} \supseteq I\left\{v \,|A_2 \cup \{v\}\right\} \tag{15}$$

Therefore, in order to illustrate the submodular of F, we can explain the equality above established. From the monotonicity of T, we can express the inequality below.

$$T\left(A_1 \cup \{v\}\right) \geq T\left(A_2 \cup \{v\}\right) \tag{16}$$

$$\Leftrightarrow I\left\{v \,|A_1 \cup \{v\}\right\} \supseteq I\left\{v \,|A_2 \cup \{v\}\right\} \tag{17}$$

Hence, F is a submodular function. For an arbitrary non-negative monotonic submodular function G, we suppose A is a solution from a greedy algorithm and B is an any other solution, the conclusion as follows is always established [11].

$$G\left(A\right) \geq \left(1 - \frac{1}{e}\right) \cdot G\left(B\right) \tag{18}$$

$$F\left(S\right) \geq \left(1 - \frac{1}{e}\right) \cdot F\left(D\right) \tag{19}$$

$$\Rightarrow T\left(S\right) \leq \left(1 - \frac{1}{e}\right) \cdot T\left(D\right) + \frac{C}{e} = \frac{1}{e}\left(C - T\left(D\right)\right) + T\left(D\right) \tag{20}$$

Introduced into our work, we can get the approximation ratio of T which means the solution obtained by the greedy algorithm has a function value at least $\left(1 - \frac{1}{e}\right)$ [12] of the maximum value of the function. The solution whose function value is at most the minimum $\left(1 - \frac{1}{e}\right)$ plus a constant term $\frac{C}{e}$ with a given C. Considering the physical meaning of the function T, the approximation guarantees the maximum approximate guarantee of the time required for the information to be spread in the network as the information source obtained by applying the greedy algorithm to solve the ITM problem.

So how should the value of C be chosen? From the above analysis, we can calculate the approximate guarantee of the function T every time a value of C is given. The value of $T(D)$ is often used as the denominator of the approximate ratio, that is the value of the shortest propagation time in the network under ideal conditions, where D is the ideal optimal set of seed nodes. When C obtains the value of $T(D)$, it can be known from equation 20 that $\frac{T(S)}{T(D)} = 1$, which is the greedy algorithm can find the optimal solution in the network, and its function value is minimum value, $T(D)$, which is the bound of the function T. However, the above situation is a virtual situation. Under realistic conditions, we do not know the value of $T(D)$. Only when we prove the approximate guarantee, we can propose such an optimal function value and obtain the approximate guarantee. $T(D)$ is actually present and has practical significance, but the process of finding it is an NP-hard problem, so we can't find it accurately at the current level of development.

5 EXPERIMENTS

We simulate a random strong-link graph that matches the actual characteristics of the Bitcoin network structure [13] [14], and the edges in the network are bidirectional. The general process of simulation is as follows. First, we select the seed set, and compare the results of the two methods (random selection and greedy strategy) in our experiments. The random selection corresponds to the traditional mode of communication. For an independent propagation process, the initial time is set to 0. In each timestamp, the process of information propagation turns to the next step, and the number of timestamps required for all nodes in the network to be infected is recorded. Due to the uncertainty of the propagation process, we conduct 200 independent experiments for each seed set, and calculate the expectation of the results to obtain the corresponding propagation time.

Figure 3 is a plot of propagation time versus number of seed nodes, where the horizontal axis is the number of seed nodes and the vertical axis is the propagation time. We experiment in a network with the size of 200 nodes. We separately select 1, 2, 3, 4, and 5 seed nodes in the same network for two methods and compare their respective required propagation times in the 5 cases. It can be seen from the figure that both results decrease as the size of the seed set increases, which is consistent with the monotonicity of the objective function mentioned in the previous section. The performance of the greedy algorithm is better than the random method for different seed numbers. According to the curves, we can also find that as the number of seeds increases, the gap between the two results decreases. We consider the limit case, when the number of seeds increases to the total number of nodes in the network, the propagation time is 0. From this we guess that when the number of seeds increases to a certain number, the time required for both methods may be an identical value. However, this situation does not match the actual situation, because the information dissemination in the network often starts from a small number of users. In addition, the routing overhead we dont consider will become the main factor in information propagation for this case.

Figure 4 shows the relationship between propagation time and network size. The horizontal axis is the network scale. In this round, we generate five different networks of size 50, 100, 150, 200, and 250. We use two methods to select seeds from the above five networks and compare their propagation time. In this case, the number of seed sets is unified to 2. According to the curve, as the network scale expands, the propagation time increases. However, for the greedy strategy, the propagation time of the 200-node network is less than the network with size of 150. It can be seen that the diversity and randomness of the network structure will make the propagation time non-monotonous. The time required by the greedy algorithm is less than the random selection method. This conclusion is consistent with the statement shown in Figure 3, which shows that the new propagation mode can greatly reduce the propagation time in Bitcoin network, which further behaves that the greedy algorithm can provide a better performance solution for the ITM problem.

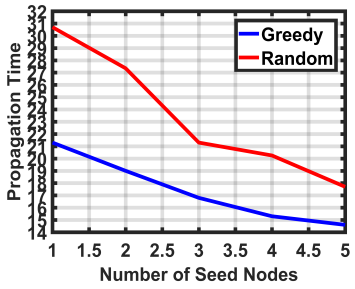

Figure 3: Relationship between Propagation Time and Number of Seed Nodes.

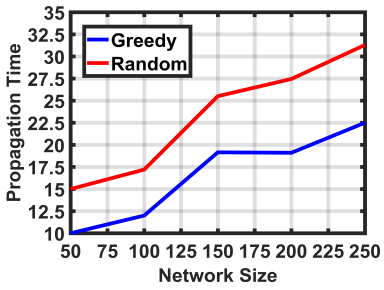

Figure 4: Relationship between Propagation Time and Network Size.

6 CONCLUSION

In this paper, we study the information propagation model in Bitcoin network. In order to solve the bifurcation problem in the blockchain system, we define a new propagation mode and propose an ITM problem to solve the selection problem of the seed node set. We first prove that the ITM problem is NP-hard and design a greedy algorithm. We prove that the objective function is monotonic and submodule. We further propose the theoretical approximate guarantee. Moreover, we discuss the upper bound of information dissemination time, which provides a theoretical basis for the selection of value C in specific problems. Finally, we verify the superiority of the greedy algorithm through experiments. By comparing with the time required for the propagation of traditional random source nodes, we find that the solution proposed by the greedy algorithm greatly reduces the time consumed for the propagation process. However, this type of communication essentially introduces a setting that violates the decentralized nature of the blockchain. But this setting is only considered from the perspective of information dissemination, which has no effect on the property of the decentralized transaction in Bitcoin network. And there are some similar situations in Bitcoin system. For example, the mine pool provides a centralized management mechanism for miners' mining and reward distribution processes. The new communication model requires that information be presented in a chronological order, which is easily met in real situations. In the future, we will further explore the selection of value C. We expect to find a precise value of C, which can be applied to the blockchain system.

REFERENCES

[1] Jordi Herrera-Joancomartí. Research and challenges on bitcoin anonymity. In *Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance*, pages 3–16. Springer, 2014.
[2] Giulia Fanti and Pramod Viswanath. Deanonymization in the bitcoin p2p network. In *Advances in Neural Information Processing Systems*, pages 1364–1373, 2017.
[3] Shaileshh Bojja Venkatakrishnan, Giulia Fanti, and Pramod Viswanath. Dandelion: Redesigning the bitcoin network for anonymity. *arXiv preprint arXiv:1701.04439*, 2017.
[4] Oguzhan Ersoy, Zhijie Ren, Zekeriya Erkin, and Reginald L Lagendijk. Information propagation on permissionless blockchains. *arXiv preprint arXiv:1712.07564*, 2017.
[5] Francisco Prieto-Castrillo, Sergii Kushch, and Juan Manuel Corchado. Distributed sequential consensus in networks: Analysis of partially connected blockchains with uncertainty. *Complexity*, 2017, 2017.
[6] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.
[7] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Annals of mathematics*, pages 439–485, 2005.
[8] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, pages 47–48. ACM, 2011.
[9] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2007.
[10] Donald M Topkis. Minimizing a submodular function on a lattice. *Operations research*, 26(2):305–321, 1978.
[11] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functionsi. *Mathematical programming*, 14(1):265–294, 1978.
[12] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.
[13] Martí Berini Sarrias. Bitcoin network simulator data explotation.
[14] P Erdős and Alfréd Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Mathematica Hungarica*, 17(3-4):359–368, 1966.